

Educational and Psychological Measurement

<http://epm.sagepub.com>

Computerizing Organizational Attitude Surveys: An Investigation of the Measurement Equivalence of a Multifaceted Job Satisfaction Measure

Karsten Mueller, Christian Liebig and Keith Hattrup

Educational and Psychological Measurement 2007; 67; 658 originally published
online May 25, 2007;

DOI: 10.1177/0013164406292084

The online version of this article can be found at:
<http://epm.sagepub.com/cgi/content/abstract/67/4/658>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Educational and Psychological Measurement* can be found at:

Email Alerts: <http://epm.sagepub.com/cgi/alerts>

Subscriptions: <http://epm.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://epm.sagepub.com/cgi/content/refs/67/4/658>

Computerizing Organizational Attitude Surveys

An Investigation of the Measurement Equivalence of a Multifaceted Job Satisfaction Measure

Karsten Mueller

Christian Liebig

University of Mannheim

Keith Hattrup

San Diego State University

Two quasi-experimental field studies were conducted to evaluate the psychometric equivalence of computerized and paper-and-pencil job satisfaction measures. The present research extends previous work in the area by providing better control of common threats to validity in quasi-experimental research on test mode effects and by evaluating a more comprehensive measurement model for job attitudes. Results of both studies demonstrated substantial equivalence of the computerized measure with the paper-and-pencil version. Implications for the practical use of computerized organizational attitude surveys are discussed.

Keywords: *measurement equivalence; survey; administration mode; attitude measure*

The use of computers in the administration of psychological surveys in organizations has become increasingly popular. Many advantages are thought to be associated with the use of computer and "online" technology for administering psychological surveys, such as enhanced user satisfaction, ease of administration and scoring, reduction of both user and transcriber errors, increased accuracy, time savings, opportunities for complex item branching and customization of the survey, increased transparency, and cost savings (e.g., Rosenfeld, Doherty, & Carroll, 1987; Rosenfeld, Doherty, Vicino, Kantor, & Greaves, 1989). Due to the growing relevance in practice, there is a need for further empirical research on the proposed benefits of computerizing organizational surveys, and in particular on the psychometric equivalence of computerized

Authors' Note: Please address correspondence to Karsten Mueller, University of Mannheim, Department of Business and Organizational Psychology, Schloss, Ehrenhof Ost, 68131 Mannheim, Germany; e-mail: karsten.mueller@psychologie.uni-mannheim.de.

surveys with their traditional paper-and-pencil counterparts (Synodinos, Papacostas, & Okimoto, 1994).

Appropriate inferences about the meaning of scores obtained on computerized organizational surveys and, by implication, any apparent advantages of computerized testing, depend on the construct validity of the scores obtained from the instruments. A particular concern that has been raised in the literature is the extent to which computer-administered psychological measures are psychometrically equivalent with their more traditional paper-and-pencil counterparts (Donovan, Drasgow, & Probst, 2000; King & Miles, 1995; Stanton, 1998). Evidence of psychometric equivalence is seen as a logical prerequisite to the interpretation of computerized test scores in terms of preexisting knowledge of the psychological constructs the tests are designed to measure.

The issue of psychometric equivalence of computerized and paper-and-pencil tests has been addressed rather extensively for cognitive ability tests (Mead & Drasgow, 1993) but much less for attitude measures (Donovan et al., 2000). The equivalence of computerized organizational attitude surveys with their traditional paper-and-pencil counterparts, in particular, still remains largely unexplored (Donovan et al., 2000). Unfortunately, most previous research on the equivalence of computerized and paper-and-pencil noncognitive measures is limited by the use of small samples or of predominantly student samples. Given that organizational attitude surveys typically address highly personal and career-relevant aspects of job experience, such as satisfaction with pay, coworkers, supervision, and working conditions, there is a clear need to examine the psychometric equivalence of computerized and paper-and-pencil job attitude measures using large samples of working adults. As Kantor (1991) noted,

Job satisfaction surveys differ from other surveys in that the results from the former may be utilized by management for personnel policy changes which may directly affect surveys participants. . . . The unique characteristics of job satisfaction instruments set them apart from attitude and opinion surveys that have been previously administered on computers. (p. 312)

Thus, some authors have pointed out that the effects of administration format may depend on the purpose and significance of the questionnaire (e.g., Ployhart, Weekley, Holtz, & Kemp, 2003; Richman, Kiesler, Weisband, & Drasgow, 1999; Wilkerson, Nagao, & Martin, 2002).

Furthermore, most empirical research on test administration mode effects relies on comparisons between mean scores observed on the different versions of the measure (e.g., Kantor, 1991; Rosenfeld et al., 1989; Synodinos et al., 1994). In this context, much attention has been paid to the influence of social desirability on scores obtained from computerized and paper-and-pencil measures (e.g., Booth-Kewley, Edwards, & Rosenfeld, 1992; Dwight & Feigelson, 2000; Lautenschlager & Flaherty, 1990; Martin & Nagao, 1989; Richman et al., 1999). However, differential social

desirable responding is only one potential threat to the comparability of results obtained with different measurement formats (van de Vijver & Poortinga, 1997). Differences in administration modes potentially affect survey results in a variety of ways, such as in different conceptual structure of constructs assessed by the survey, differences in the slopes and intercepts relating latent constructs with item responses, or differences in the means and variances of latent attitude constructs (Byrne, Shavelson, & Muthen, 1989; Drasgow, 1984; Drasgow & Kanfer, 1985; Horn & McArdle, 1992; Reise, Widaman, & Pugh, 1993; Riordan & Vandenberg, 1994; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000).

Modern statistical approaches, such as item response theory (IRT) or confirmatory factor analysis (CFA), allow a much more rigorous comparison of the psychometric characteristics of operational measures; however, they have only rarely been used to evaluate the psychometric equivalence of paper-and-pencil versus computer-administered job attitude measures. Hence, the present research is designed to extend previous research in this area by examining in two quasi-experimental field studies the psychometric equivalence of a multifaceted measure of job satisfaction. In particular, the present research uses confirmatory factor analytic methods to examine the equivalence of a full-length job attitude measure that contains items that address satisfaction with a variety of job facets in stratified samples of employees of a large organization.

Computerized Administration of Noncognitive Measures and Measures of Job Attitudes or Perceptions

Concerns about the effects of test administration on scores have been the focus of much empirical research. As noted, most previous research on the comparability of computerized and paper-and-pencil administration of noncognitive measures, such as measures of personality, beliefs, or attitudes, has examined differences in the mean scores obtained from the different administration modes. For example, several studies reported higher levels of self-disclosure and candor in responses on computer-administered questionnaires compared to paper-and pencil versions of the measures (e.g., R. N. Davis, 1999; Turner et al., 1998). These findings are contradicted, however, by other studies demonstrating more self-disclosure or candor in the paper-and-pencil condition (e.g., Schuldt, 1988) or no differences across modes (e.g., Erdman, Klein, & Greist, 1983; Knapp & Kirk, 2003; Locke & Gilbert, 1995; Millstein, 1987).

Much of the interest in this context has been on the effects of test administration mode on socially desirable responding, which would typically have the overall effect of changing the mean level scores obtained from different test administration modes (e.g., C. Davis & Cowles, 1989; Lautenschlager & Flaherty, 1990; Potosky & Bobko, 1997; Richman et al., 1999). Thus, a number of studies have also been conducted to evaluate response distortion directly by comparing scores on direct measures of social desirability across administration modes. Taken together, the results of these studies have been equivocal (e.g., Booth-Kewley et al., 1992; C. Davis &

Cowles, 1989; Fox & Schwartz, 2002; Joinson, 1999; Kiesler & Sproull, 1986; King & Miles, 1995; Lautenschlager & Flaherty, 1990; Martin & Nagao, 1989; Potosky & Bobko, 1997; Wilkerson et al., 2002).

Furthermore, of the studies included in the meta-analysis by Richman et al. (1999) examining differences across computerized and paper-and-pencil formats, 17 studies explicitly employed direct measures of response distortion. However, of these 17 studies, 15 were based on undergraduate samples, whereas the remaining 2 were conducted in a Navy setting. Therefore, few conclusions can be drawn from this research about the likely equivalence or nonequivalence of organizational attitude surveys across administration modes under applied field conditions (Kantor, 1991; Synodinos et al., 1994).

Although studies of the equivalence of noncognitive measures across administration modes have begun to appear in the literature, very little empirical work has examined computerized measures of organizational attitudes, despite their widespread use in practice. Several studies have examined the effects of administration mode on mean-level responses on organizational attitude or perception measures. Booth-Kewley et al. (1992), for example, surveyed male Navy recruits and found no mean differences across computerized versus paper-and-pencil administration modes on Mowday, Steers, and Porter's (1979) Organizational Commitment Questionnaire. Synodinos et al. (1994) reported small differences across administration modes in responses of airport patrons regarding their satisfaction with various aspects of different airport facilities. Rosenfeld et al. (1991), using a sample of 72 undergraduates, found higher ratings on a computerized version of the Job Descriptive Index (JDI; Smith, Kendall, & Hulin, 1969) for students high in self-monitoring, but higher ratings on a paper-and-pencil condition for low self-monitors. Kantor (1991) also explicitly addressed the comparability of paper-and-pencil and computer-based versions of a job satisfaction measure, reporting no statistically significant differences in responses of civilian employees of the U.S. Navy on the five facet scales of the JDI across administration modes. Thus, as with previous research of the effects of administration mode on mean level responses of other noncognitive measures, studies of the effects of computerized versus paper-and-pencil administration of organizationally relevant attitudes and perceptions have yielded inconclusive findings overall.

More important, comparisons of mean scores across administration modes found several potential sources of psychometric equivalence or nonequivalence across modes. In particular, observed mean scores are the result of factor mean levels but also of the mean and covariance structures that link the factors with item responses. Although a number of studies have examined correlations between scores obtained from paper-and-pencil versus computerized administrations, with moderate to high correlations typically being reported (e.g. Katz & Dalby, 1981; Lushene, O'Neil, & Dunn, 1974; Potosky & Bobko, 1997; Vansickle, Kimmel, & Kapes, 1989; White, Clements, & Fowler, 1985; Wilson, Genco, & Yager, 1985), correlations between test modes also reflect the covariance structures that underlie item responses. Hence,

there is a need for systematic field research designed to examine the psychometric equivalence of computerized versus paper-and-pencil administrations of job attitude measures using analytic procedures that provide more thorough insight about the potential sources of equivalence and nonequivalence of scores obtained from different administrations of a common measure.

Analysis of Measurement Equivalence

Despite the growing awareness of issues regarding measurement equivalence, only a handful of studies have explicitly compared the measurement equivalence of organizational surveys across paper-and-pencil versus computerized administration modes. In a sample of undergraduate business majors, King and Miles (1995) observed equivalent measurement in analyses using CFA across modes for several scales designed to assess social desirability, self-esteem, equity perception, and Machiavellianism. Evidence of greater impression management was observed for the paper-and-pencil version, which the authors concluded was the result of true differences and not an artifact of measurement nonequivalence. In a study focused more specifically on the measurement of organizational attitudes, Stanton (1998) examined the equivalence of an instrument assessing perceptions of supervisors across paper-and-pencil versus computerized administration modes. Using a sample of 231 professional employees from several organizations, scores on a measure of organizational justice developed by Niehoff and Moorman (1993) exhibited equivalent factor structures, measurement weights, and covariances across modes in analyses using multiple group CFAs (MGCFAs).

Only one empirical study has specifically addressed the equivalence of job satisfaction measures across administration modes. Using IRT, Donovan et al. (2000) examined the cross-mode measurement equivalence of two facets of the JDI, namely, satisfaction with leadership and with coworkers. Tests of measurement equivalence were based on comparisons across three organizations including a large public utility company, a large food-processing company, and a midwestern university, in which either a paper-and-pencil version or computerized version of the measures were administered. IRT analysis for the satisfaction with supervisor scale were based on a comparison of the university sample completing the computerized version of the survey with the participants from the food-processing company who were administered a paper-and-pencil version of the JDI scale. Results showed no evidence of differential test functioning (DTF) across media. Item level analyses, which involved visual inspection of the item response functions (IRFs), also indicated equivalent measurement across the two samples completing the computerized or paper-and-pencil version of the measure. In the same study, analysis of the satisfaction with coworkers scale compared the university sample completing a computerized version with participants of the food-processing company and the public utility company who were administered the measure in paper-and-pencil format. These

analyses showed that although there was no evidence of DTF in the comparison of the university sample completing the computerized version and the public utility completing the paper-and-pencil version, the coworker scale showed DTF when the university sample was compared with the food-processing sample. Examination of the IRFs revealed that two items functioned differentially across these groups.

Because the computerized and paper-and-pencil versions of the survey were administered in different organizations to workers performing very different jobs, Donovan et al. (2000) noted that differences in IRFs might be due to sample differences rather than differences in administration mode. Consequently, the confounding of environment (e.g., organization and job) with mode of administration threatens the validity of inferences about test mode effects observed in their study (Shadish, Cook, & Campbell, 2002). Thus, the authors concluded that differences in the computerized and the paper-and-pencil version of the instruments could not be unambiguously attributed to either valid group differences or to administration mode differences. The establishment of measurement equivalence across modes in comparisons involving the satisfaction with supervisor scale, or in the comparison involving university respondents with the public utility sample, does not render the problem of nonequivalent samples irrelevant. Using designs that confound sample characteristics with administration modes carries the risk that existing differences across the application modes remain undetected because they are compensated for by differences in sample characteristics, resulting in invalid inferences of measurement equivalence across modes. Donovan et al. (2000), therefore, appropriately recommended that future research should obtain samples from within a single organization to control for any potential confounding effects of environment on conclusions about measurement equivalence across computerized and paper-and-pencil version of the same job satisfaction measure.

The study by Donovan et al. (2000) and their conclusions about the ambiguous nature of the results given the nonrandom sampling used in their study highlight problems commonly encountered in this field of research. Researchers who wish to address the invariance of job satisfaction measures in applied field settings using samples of current organizational members rather than student samples most often have to rely on some type of quasi-experimental design (Shadish et al., 2002). Quasi-experiments, of course, vary in the degree in which they allow control over threats to internal validity and construct validity, in addition to external validity and statistical conclusion validity (Cook & Campbell, 1979; Shadish et al., 2002). Quasi-experiments may permit valid inferences about causal relationships when they control for alternative explanations of observed differences. Selection is perhaps the most salient threat to the validity of inferences about causal relationships arising from field studies with non-randomly sampled participants. Thus, field studies that examine measurement equivalence across computerized and paper-and-pencil test modes must be designed to minimize the likelihood that observed differences across modes are caused by differences in sample characteristics. This is particularly true

given that workers in different organizations who perform different jobs may differ in a number of ways. Hence, there is a clear need for carefully conducted empirical research of test mode effects using designs for field settings that minimize threats to validity, such as selection. As Donovan et al. (2000) noted, "It is recommended that future investigators try to obtain two or more samples from the same organization (or very similar organizations) for an unconfounded assessment of DTF" (p. 312).

Given the above literature, the present study contributes in two significant ways. First, the equivalence of computerized and paper-and-pencil versions of a job satisfaction measure is examined using data collected in carefully designed field research that addresses limitations of previous studies. In particular, all participants in the present research were sampled from within the same organization, performed equivalent tasks, and were stratified for the analysis in regards to available job-related characteristics. Second, building on previous research, the job satisfaction measure used in the present study expands the examination of measurement equivalence across application modes by including a comprehensive hierarchical model for job satisfaction that includes additional job facets such as satisfaction with working conditions, the work itself, promotion, company and management, flow of information, innovation, cooperation with coworkers, and supervision. Measurement equivalence is examined utilizing an MGCFA framework. In Study 1, the attitude scales were administered during a large-scale employee attitude survey. Participants of a specific field site within the company were administered a computerized version of the job satisfaction measure, whereas all other employees working at 11 other field sites were administered a paper-and-pencil version of the survey. In Study 2, customer service employees working at various customer sites of the same organization that participated in Study 1 completed the attitude surveys using either a paper-and-pencil version or the computerized version.

Study 1

Method

Participants. The sample included 6,112 employees of a large multinational manufacturer of industrial facilities who participated in a full-scale employee attitude survey designed to assess job satisfaction among virtually all members of the organization. Participants were all members of the same organization but were located at 12 different field sites geographically spread across the United States of America. One field site was selected for the computerized administration of the survey ($N = 2,529$), whereas employees of all other 11 field sites received a paper-and-pencil version of the survey ($N = 3,583$). For each employee participating in the survey, information was received about their employment classification (hourly, exempt, nonexempt, management) and on whether they performed a supervisory function. To control for

Table 1
Numbers of Participants Online and Offline Sample

	Stationary Field Sites		Customer Service Employees	
	Offline (N = 1,696)	Online (N = 1,696)	Offline (N = 239)	Online (N = 239)
Employee classification				
Hourly	56	56	16	16
Nonexempt	537	537	62	62
Exempt	798	798	98	98
Management	305	305	63	63
Supervisor function				
Supervisor function	632	632	128	111
Nonsupervisor function	1,064	1,064	239	239

sample differences between the computerized and paper-and-pencil conditions, we stratified the samples on the basis of employee classification and supervisory function, using random selection procedures to arrive at an equal number of participants in each classification and status across the groups participating in the study. This procedure resulted in a sample size of 1,696 for each condition. Table 1 displays the distribution of participants in each category in each of the two groups for Study 1 and Study 2.

Measure. The multi-item faceted job satisfaction measure that was used in this research was developed several years ago by an international research and consulting institute in Germany to assist organizations in the evaluation of employee attitudes. The measure has been translated into many different languages and has been used in many international applications. The measure used in this research contained 31 items assessing classical job satisfaction facets such as working conditions, the work itself, promotion, company and management, flow of information, innovation, coworkers, and supervision. Satisfaction with compensation was not evaluated in the present organization due to concerns expressed by the key members of the organization. Between three and five items with 5-point Likert scales were written for each facet. Both the computerized and paper-and-pencil versions of the survey contained exactly the same items, with the same 5-point Likert response scale.

To evaluate the construct validity of scores obtained from the measure, confirmatory factor analyses were performed using samples of employees who had been administered the paper-and-pencil version of the instrument in previous years in the same organization. Analyses with AMOS 5.0 were performed in two samples, including 14,975 employees who were administered the measure in 1999 and 15,141 employees who responded in 2000. For these analyses, a model was tested that estimated paths linking items to their hypothesized facets and linking the facets to a single higher order factor.

The model did not allow for any correlated errors. The measurement weights of a single item for each facet was fixed to 1.0 to set the scale of the first-order factors, and a single path linking one of the facets to the general factor was set to 1.0 to set the scale of the higher order factor. Estimation was performed using maximum likelihood. The comparative fit index (CFI; Bentler, 1988) was greater than .94 in each sample, and the root mean square error of approximation (RMSEA; Steiger, 1990) was .047 and .048 in the 1999 and 2000 samples, respectively, indicating good fit of the hypothesized hierarchical model underlying the job satisfaction measure. Internal consistency reliability (Cronbach's alpha) ranged from .785 to .920 in the 1999 sample and .795 to .925 in the 2000 sample. For the actual data used in the following studies, internal consistency coefficients ranged from .848 to .928 in Study 1 and .863 to .931 in Study 2. These results suggest that items were highly correlated within the facets (Henson, 2001).

Convergent validity of scores from the measure used in this research with other published job attitude measures was evaluated in a separate sample of 100 employees in Germany from a variety of occupations who completed the measure along with German translations of the JDI and the Minnesota Satisfaction Questionnaire (MSQ) Short-Form (Weiss, Dawis, England, & Lofquist, 1967). The overall satisfaction composite score formed from items in the present measure correlated .868 with the overall satisfaction score of the JDI and .833 with the overall score of the MSQ, indicating strong convergent validity of scores from the present measure with other scores from other well-known instruments. Furthermore, correlations of the facets of the present measure with corresponding scales of the JDI (work itself, promotion, supervision, and coworkers) ranged from .484 to .842. Taken together, analyses of the construct validity of scores from the present measure showed good support for the hypothesized factor structure of the instrument, strong internal consistency, and excellent convergence with scores from other better known job satisfaction instruments.

Procedure. As noted above, the present study was conducted in the context of a full-scale employee opinion survey. Participation in the survey was voluntary in both the paper-and-pencil and computer-administered conditions; no positive or negative sanctions resulted from participation or nonparticipation. As mentioned above, 1 site was chosen to receive the computerized version of the survey, whereas the other 11 received the paper-and-pencil version. The field site that was chosen to participate in the computerized administration was selected primarily because the internal survey project group that served as a liaison between the organization and the external consulting institute was located at this site, and therefore the accessibility and quality assurance of the e-mail-address databases needed for the computerized administration were most easily managed at this site. Participants at this site were sent an e-mail containing an invitation to participate and complete instructions about the computer administration, such as logging procedure, survey completion, and data submission. In particular, the e-mail included a unique access code for each participant that allowed computer access to the survey using the company's intranet. Each access code was valid

only once to ensure that no employee completed the online survey more than once. To ensure anonymity of the research participants, two unique databases were developed. The first contained e-mail addresses and the login codes, and the second contained the login codes and the survey items themselves. After completing the survey (i.e., one participant answered all questions), the corresponding access code for a participant expired. When the period of data collection ended, the first database containing the link between access code and the e-mail address was erased.

The computerized version of the questionnaire was administered in a WWW environment using HTML forms and PHP databases because of the widespread familiarity of organization members with Internet browsers. The computerized version was divided into several pages. Each page was transmitted to the server separately. Respondents were not able to backtrack or edit previous items once the corresponding page was sent to the server. If the participant logged off or the connection was terminated for some reason, the participant could continue the survey after a renewed login. Within each page, the respondent could modify his or her answers by clicking on another alternative. If an item was left blank the participant was reminded just once; if it was left blank again, the missing value was accepted. During the completion of the questionnaire, a bar at the top of each page indicated the status of completion.

Participants in the paper-and-pencil administration condition of the survey received their questionnaire in a sealed envelope. Responses to items were given on a separate answer sheet, and this was put in a self-addressed envelope which could be dropped off in several "mailboxes" around the work site. Envelopes, which each contained one answer sheet, were then sent to the consulting institute, where the data were recorded electronically. The preliminary remarks and instructions about the purpose of the survey were identical in all field sites, regardless of administration mode.

Analyses. We assessed measurement equivalence across the computerized and paper-and-pencil versions of the measure in several steps using an MGCFA framework. In general, existence of measurement equivalence is tested by a series of nested models, evaluating the appropriateness of cross-group equality constraints based on the change in model fit indices due to the imposed constraints. As a first step, configural invariance was tested by evaluating the equivalence of the basic model forms across groups. To assess equivalence of model form, we evaluated the fit of a hierarchical factor model (e.g., Byrne et al., 1989) that specified a single second-order factor with structural path to each of the eight facets (working conditions, work itself, promotion, company and management, flow of information, innovation, coworkers, and supervision), which were linked to specific items written for each respective factor. The measurement weights and intercept of one item for each of the first-order factors, and one path and latent intercept linking the first-order factor with the second-order factors, were set to 1.0 to set the scaling of the latent facet and general factors (Bollen, 1989; Kline, 1998). All other model parameters were freely estimated using maximum likelihood estimation with AMOS 5.0.

The chi-square index of overall model fit is criticized for being overly sensitive to minor deviations from the proposed model, especially in large sample studies (Bollen, 1989; Bollen & Long, 1993), as in the present research. Thus, several authors recommend the use of practical fit indices to evaluate overall model fit (Bollen, 1989; Bollen & Long, 1993; Kline, 1998), including the confirmatory fit index (CFI; Bentler, 1988), the incremental fit index (IFI; Bentler & Bonett, 1980) and the root mean square error of approximation (RMSEA; Steiger, 1990), which were examined in the present research. Values of .90 or above for the CFI and IFI, and values of .08 or lower for the RMSEA, are usually taken as evidence of adequate fit (e.g., Vandenberg & Lance, 2000). Hu and Bentler (1999), however, recommended using cutoff values close to .95 for the CFI and IFI and values close to .06 for the RMSEA to indicate good model fit. Several authors, however, have noted that the recommendations of Hu and Bentler may be premature (e.g., Dieffendorff, Silverman, & Greguras, 2005; Vandenberg & Lance, 2000). Thus, in the present study, the commonly accepted cutoff values (CFI, IFI \geq .90 and RMSEA \leq .08) are to be considered as the lower bound for good fit, whereas the cutoff values suggested by Hu and Bentler indicate excellent fit.

As a second step in evaluating equivalence across the two administration modes, metric equivalence was tested in the MGCFA framework. Equivalence of the measurement weights implies that the slopes relating observed item scores with their respective latent factor are equivalent across groups, and thus, the rating scales are used equivalently across groups (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). Poorer overall fit associated with models that impose between-group equality constraints compared to free models suggests that the equality constraints are untenable.

In a third step, the scalar invariance (Meredith, 1993; Steenkamp & Baumgartner, 1998) is tested by evaluating the equivalence across groups of the intercepts linking items with factors. Empirical evidence of scalar invariance implies that differences in observed item or scale means are only due to differences in latent factor means (Cheung & Rensvold, 1999). To test for scalar equivalence, the intercepts linking items with factors are constrained to be equal, and the decrease in fit compared to a model that freely estimates these intercepts is used to determine whether the hypothesis of equivalent intercepts is tenable.

Finally, latent mean differences are tested by further constraining the latent means to equality across groups. Analyses of latent means are powerful in testing substantive hypothesis about mean differences because they are based on more reliable estimates of the true underlying constructs of interest. For example, format-based differences in social desirable responding affecting each item of a facet to more or less the same degree would be reflected in differences at the construct level (Little, 2000).

Often, researchers have examined the change in chi-square between nested models when the interest has been in evaluating the effects of imposing equality constraints between groups (e.g., Bollen, 1989; Widaman, 1985). When sample sizes are large, statistically significant values of the chi-square difference statistic may be observed, even when differences in parameter estimates are trivial (e.g., Cheung & Rensvold,

Table 2
Results of Multiple Group Analyses of Model
Parameter for Stationary Field Personnel

Model	χ^2	df	CFI	IFI	RMSEA
Equivalent form	5,400.212	852	.947	.947	.040
Equivalent first-order measurement weight	5,475.183	875	.946	.946	.039
Equivalent intercepts	5,575.751	898	.945	.945	.039
Equivalence of latent means	5,704.677	906	.944	.944	.040

Note: CFI = confirmatory fit index; IFI = incremental fit index; RMSEA = root mean square error of approximation.

2002; see also Ryan, Chan, Ployhart, & Slade, 1999). Cheung and Rensvold (2002), therefore, recommended that researchers examine changes in the values of practical fit indices when the interest is in evaluating the significance of differences between nested models. Based on a large-scale Monte Carlo investigation, Cheung and Rensvold concluded that changes of the CFI (Δ CFI) that are less than or equal to .01 imply no substantially significant difference between nested models. Values of the Δ CFI that equal to .01 suggest that the constraints should be viewed with caution, whereas values of Δ CFI that are greater than .01 indicate practically significant between-group differences in the constrained parameters.

Results

Results of the tests of configural equivalence for the samples completing the computerized and paper-and-pencil versions of the measure are presented in Table 2. Table 2 also presents the results for the test of metric equivalence that postulates equal measurement weights across media, as well as the results of tests of scalar invariance and latent mean differences. As can be seen in Table 2, the proposed baseline model for the attitude survey fits quite well across the samples completing the two versions of the survey. The CFI and IFI were both .947 and the RMSEA was below .05 for the model that imposed constraints only on the form of the model across the groups. In other words, the patterns of significant and nonsignificant measurement weights are equivalent across the different versions of the instrument, as indicated by the good fit of the baseline model. As can be seen in Table 2, restrictions on the equivalence of the measurement weights also showed good support for metric equivalence across administration modes. Restricting the item measurement weights to equality across groups resulted in a decrease of .001 in the CFI and IFI, suggesting trivial differences in item measurement weights across media. The RMSEA showed slightly *better* fit of this more constrained model, further supporting the conclusion that item scalings were equivalent across the computerized and paper-and-pencil versions of the instrument.

The test for scalar invariance also showed a small decrease in the fit of the model compared to when measurement weights were constrained to equality across groups. Results show that constraining the intercepts to equality across media resulted in a drop in the CFI of .001 and no change in the RMSEA. Further constraining the latent means to equality across media led also to a drop in the CFI of .001, and an increase of only .001 in the RMSEA, indicating that differences in the means on the latent constructs were not practically significant across survey administration mode.

Discussion

Results of Study 1 provide good support for the hypothesis that job satisfaction can be measured equivalently across paper-and-pencil and computerized administrations. Furthermore, analysis of latent means and intercepts showed that observed scores will not only reflect the same construct across administration modes, but scale scores will also exhibit the same raw score level. The present study is unique in that it provides a more thorough assessment of the measurement equivalence of a job attitude measure and, in particular, a large-scale faceted job satisfaction measure that includes most of the important facets discussed in the literature, across administration modes. Moreover, compared to previous research in this area (e.g., Donovan et al., 2000), the present study provides more control over common threats to validity, particularly selection, and thus provides a stronger basis for inferences about the equivalence or nonequivalence of job satisfaction measures when administered in computerized versus paper-and-pencil formats. Because participants in the two administration conditions were all members of the same organization, it is less likely that results are substantially complicated by sample differences compared to the research reported by Donovan et al. (2000). Nevertheless, the fact that participants in the two conditions of this study were employed in different field sites, albeit of a single parent company, implies that there may be some unknown and, therefore, uncontrolled, differences between the two samples. For that reason, we conducted a second field study to evaluate measurement equivalence across modes, using samples of participants who were drawn from across a large number of field sites of the same organization.

Study 2

Method

Participants and procedure. In addition to the 6,112 members of the organization who participated in Study 1, 760 customer service field representatives also completed the same job satisfaction measure either by computer or paper and pencil. Participants in Study 2 were employed by the same organization as the sample used in Study 1 but were located at customer sites rather than any specific field site of the

parent organization. Of these 760 participants, 373 completed the computerized version of the questionnaire, whereas the other 387 completed the paper-and-pencil version. Those who completed the computerized version were formally attached to the same location that administered the computerized measure in Study 1; but this time, participants were not physically located at the same field site, and hence, their socialization was probably determined as much by their client organizations as by the parent organization. Participants who completed the paper-and-pencil measure in Study 2 were formally attached to a variety of specific field sites of the parent organization; however, again, they were physically located at a client's site and interacted regularly with members of the client organization. Information about each participant's employment classification and status was available, as in Study 1. To control differences across samples, the samples were stratified as in Study 1. The resulting sample size was 239 in each condition (Table 1). The procedures used for administering the measures were the same as in Study 1. Participants who completed the computer-administered measure were sent an e-mail requesting their participation and explaining the procedure for logging in and entering their responses, just as in Study 1. Likewise, participants who completed the paper-and-pencil version were sent the questionnaire by interoffice mail and then placed their completed questionnaires in self-addressed envelopes that were placed in interoffice mail.

Measures and analysis. Study 2 used exactly the same instrument that was used in Study 1. The same 31 items were analyzed with AMOS 5.0 to evaluate measurement equivalence across the two administration modes using MGCFA. As in Study 1, the first step in these analyses evaluated equivalence of model form, and subsequent steps imposed restrictions on the equality of item measurement weights, item intercepts, and latent means, respectively. The changes in CFI that resulted from imposing constraints on the equivalence of model parameters were used to evaluate whether administration mode affected item measurement weights, intercepts, and factor means.

Results and Discussion

Results of tests of the equivalence of the measure across administration modes for the samples of customer service employees are presented in Table 3. As can be seen in the table, the baseline model that constrained only model form to equivalence across groups fit the data very well, as indicated by the CFI and IFI values and the RMSEA. Constraining item measurement weights to equivalence across groups led to a drop in the CFI of .01, suggesting only trivial differences in these parameters across groups. Tests for scalar invariance that constrained item intercepts to equality across formats resulted in an increase in the CFI of .01. This indicates that the increase in degrees of freedom due to constraining the intercepts across groups overcompensated for the trivial increase in the model discrepancy. Finally, the constraining the latent means to equivalence across administration modes resulted in a

Table 3
Results of Multiple Group Analyses of Model
Parameter for Customer Service Personnel

Model	χ^2	<i>df</i>	CFI	IFI	RMSEA
Equivalent form	1,756.347	852	.924	.924	.047
Equivalent first-order measurement weights	1,803.308	875	.921	.922	.047
Equivalent intercepts	1,822.286	898	.922	.922	.047
Equivalence of latent means	1,839.938	906	.921	.921	.047

Note: CFI = confirmatory fit index; IFI = incremental fit index; RMSEA = root mean square error of approximation.

decrease in the CFI of .001, again indicating no substantial differences in the latent means across the paper-and-pencil and computerized administrations of the measure. Thus, as with the comparisons involving employees of different field sites who completed different versions of the same measure in Study 1, comparisons across the customer service representatives who were dispersed across a large number of customer sites indicated psychometric equivalence of the job satisfaction measure when administered by computer and by a more traditional paper-and-pencil format.

General Discussion

The purpose of this research was to extend existing research on the psychometric equivalence of paper-and-pencil and computerized job attitude measures administered in field settings to actual organizational members. In particular, the present research sought to minimize two major limitations of previous research. First, we examined the equivalence of a multifaceted job satisfaction measure across administration modes with data collected under applied field conditions. This research setting maximizes the findings' generalizability to the actual intended application of the instrument. As noted above, only one previous empirical study has investigated the psychometric equivalence of computerized and paper-and-pencil job satisfaction measures (Donovan et al., 2000), and this study confounded administration mode differences with differences in the organizations and jobs of participants who completed either the computerized or paper-and-pencil instruments. In contrast, the present study minimized problems with selection as a threat to the internal validity of comparisons across test administration modes. Participants in the present research who completed different versions of the measure were members of the same organization, performing similar tasks, and were stratified according to available job related characteristics. Furthermore, this research included two separate studies of test administration mode effects that were each designed to address selection as a threat to the validity of inferences in unique ways.

Second, the present research extends previous work in this area by including a broader array of job satisfaction facets in the comparisons between test administration modes. Previous work has examined the psychometric equivalence of computerized versus paper-and-pencil measures of satisfaction with coworkers and with supervisors but has not examined the equivalence of measures of a variety of other facets that are widely believed to underlie job attitudes. The present research, in contrast, examined the equivalence of a comprehensive hierarchical model for job satisfaction across test administration modes.

Using a CFA framework to test for measurement invariance, we found strong support for the equivalence of the job satisfaction measure across paper-and-pencil and computerized test administration modes. We found no evidence of an effect of the mode of test administration on the factor structure of the questionnaire or for the appropriateness of individual items to serve as indicators of hypothesized factors. Overall, previous research on the equivalence of computerized and paper-and-pencil measures of noncognitive constructs has been equivocal (Richman et al., 1999), with about as many results showing differences across modes and results showing that the two administration modes are psychometrically equivalent. We also found evidence of equivalence in the intercepts linking paper-and-pencil versus computerized job attitude items with the latent job satisfaction facets, implying that differences in observed scale means between modes are determined exclusively by differences in the latent job attitude constructs. These findings suggest that previously reported differences in measurement characteristics of job satisfaction measures across administration modes were most likely caused by confounding differences in sample characteristics across groups. Furthermore, the consistency of the results across the two studies reported here supports the generalizability of our overall conclusions about the equivalence of job satisfaction measures across administration modes.

Implications for Practice

The results reported here are relevant for practice. As noted above, the use of computers for the administration of psychological measures is increasing in organizational practice, for example, in the selection of job applicants (e.g., Burke, 1992; Ployhart et al., 2003; Potosky & Bobko, 1997) or in the evaluation of organizationally relevant perceptions or attitudes (e.g., Bartram & Bayliss, 1984; Booth-Kewley et al., 1992; Synodinos & Brennan, 1990). In evaluating organizationally relevant attitudes, it is important to establish measurement equivalence because a large number of interventions might be derived from comparisons of data collected in the context of company-internal benchmarking. In such contexts, it may not be practically feasible to administer measures using a single administration mode because of differences across organizational units or individual respondents in the availability or quality of computer and intranet resources. Similarly, measurement equivalence would be necessary if comparisons are desired across time between results obtained

from computerized administrations with previous results obtained through the administration of paper-and-pencil measures. External benchmarking is also common, whereby results are compared across organizations that collect comparable data relevant to organizational attitudes and perceptions. Thus, the design of appropriate organizational interventions requires comparability across time or organizational units, or between different organizations despite whether data are collected using traditional paper-and-pencil measures or more modern computerized media. The results of the present research strongly suggest that scores obtained from paper-and-pencil and computerized administered job satisfaction measures can be compared with confidence.

Limitations and Suggestions for Further Research

The findings of the present research must be interpreted in the context of a few potential limitations. First, we did not use purely random assignment of research participants to test administration conditions. Randomization is critical for supporting the internal validity of empirical findings and is an important concern in quasi-experimental field research. Although the present research eliminated several sources of nonrandom assignment of participants to conditions (i.e., all participants were employed by a single organization; samples were stratified on the basis of available job-related characteristics; and in Study 2, participants were recruited from several different field sites), there might have still been some uncontrolled differences between participants who completed the paper-and-pencil versus the computerized measures. Although there are no obvious confounding characteristics between participants in the present research who completed the different versions of our job satisfaction measure, some differences may still have existed, of course. Clearly, the present research design is a substantial improvement over previous work in this area (Donovan et al., 2000) in terms of both the scope of our investigation, particularly with regards to the breadth of job satisfaction facets that were investigated, and in the control over selection as a threat to internal validity.

Second, the present findings may be specific to the context of administration, particularly with regards to issues of anonymity or personal significance of the questionnaire purpose. Some authors, for example, have suggested that these factors may influence the results of computerizing psychological measures compared to administering the same measures using a traditional paper-and-pencil approach (e.g., Booth-Kewley et al., 1992; Fox & Schwartz, 2002; Joinson, 1999; Richman et al., 1999; Wilkerson et al., 2002). In the present study, participants were assured that their responses would be anonymous, and the procedures used in administering both versions of the measure were clearly designed to ensure that anonymity. Future research is needed that examines the moderating effects of anonymity and questionnaire purpose on the psychometric equivalence of computerized organizational attitude measures with their paper-and-pencil counterparts.

Finally, we did not examine the measurement equivalence across administration modes of a well-known measure of job satisfaction, such as the JDI or MSQ, but instead examined a newer measure that was developed by a German research and consulting institute for use by a variety of organizations, including multinational corporations. Hence, the generalizability of our results to the more popular measures may be a concern. Nevertheless, we used a variety of analytic approaches to demonstrate the construct validity of scores from the present instrument and its convergent validity with scores both from the JDI and MSQ. Results of these analyses strongly suggest that our findings are unlikely to be unique to the particular measure examined in this research.

Despite the few limitations noted above, the results provide strong support that paper-and-pencil job satisfaction measures can be transformed for computerized administration without compromising the validity of inferences rendered about the meaning of test scores. No differences were observed in the measurement characteristics of the instrument across administration modes or in the means of the measure or latent job attitude constructs. Hence, researchers and practitioners may be able to use computerized measures to evaluate job attitudes, and scores on these measures may be compared with those obtained from paper-and-pencil measures to provide insights of a diagnostic nature across organizational units, across time, or between different organizations, even when data are collected using different modes of administering the job attitude items. Clearly, additional research is needed to evaluate whether there are important moderators of the effects examined in the present research, and we encourage researchers to pursue such research questions using carefully designed quasi-experimental investigations in field settings.

References

- Bartram, D., & Bayliss, R. (1984). Automated testings: Past, present, and future. *Journal of Occupational Psychology*, 57, 221-237.
- Bentler, P. M. (1988). Causal modeling via structural equation systems. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology. Perspectives on individual differences* (2nd ed., pp. 317-335). New York: Plenum.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bollen, K. A. (1989). *Structural equations with latent variables*. Oxford, UK: John Wiley.
- Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.
- Booth-Kewley, S., Edwards, J. E., & Rosenfeld, P. (1992). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology*, 77, 562-566.
- Burke, M. J. (1992). Computerized psychological testing: Impacts on measuring predictor constructs and future job behavior. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 203-238). San Francisco: Jossey-Bass.
- Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures. The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.

- Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, 25, 1-27.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- Cook, T. D., & Campbell, T. L. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton-Mifflin.
- Davis, C., & Cowles, M. (1989). Automated psychological testing: Method of administration, need for approval, and measures of anxiety. *Educational and Psychological Measurement*, 49, 311-320.
- Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments & Computers*, 31, 572-577.
- Dieffendorff, J. M., Silverman, S. B., & Greguras, G. J. (2005). Measurement equivalence and multisource ratings for non-managerial positions: Recommendations for research and practice. *Journal of Business and Psychology*, 19, 399-425.
- Donovan, M. A., Drasgow, F., & Probst, T. M. (2000). Does computerizing paper-and-pencil job attitude scales make a difference? New IRT analyses offer insight. *Journal of Applied Psychology*, 85, 305-313.
- Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin*, 95, 134-135.
- Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70, 662-680.
- Dwight, S. A., & Feigelson, M. E. (2000). A quantitative review of the effect of computerized testing on the measurement of social desirability. *Educational and Psychological Measurement*, 60, 340-360.
- Erdman, H., Klein, M. H., & Greist, J. H. (1983). The reliability of a computer interview for drug-use abuse information. *Behavior Research Methods & Instrumentation*, 15, 66-68.
- Fox, S., & Schwartz, D. (2002). Social desirability and controllability in computerized and paper-and-pencil personality questionnaires. *Computers in Human Behavior*, 18, 389-410.
- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement & Evaluation in Counseling & Development*, 34, 177-189.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments & Computers*, 31, 433-438.
- Kantor, J. (1991). The effects of computer administration and identification on the job descriptive index (JDI). *Journal of Business and Psychology*, 5, 309-323.
- Katz, L., & Dalby, J. T. (1981). Computer and manual administration of the Eysenck Personality Inventory. *Journal of Clinical Psychology*, 37, 586-588.
- Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly*, 50, 402-413.
- King, W. C., & Miles, E. W. (1995). A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. *Journal of Applied Psychology*, 80, 643-651.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.
- Knapp, H., & Kirk, S. A. (2003). Using pencil and paper, Internet and touch-tone phones for self-administered surveys: Does methodology matter? *Computers in Human Behavior*, 19, 117-134.
- Lautenschlager, G. J., & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology*, 75, 310-314.
- Little, T. D. (2000). On the comparability of constructs in cross-cultural research. A critique of Cheung and Rensvold. *Journal of Cross-Cultural Psychology*, 31, 213-219.

- Locke, S. D., & Gilbert, B. O. (1995). Method of psychological-assessment, self-disclosure, and experiential differences. A study of computer, questionnaire, and interview assessment formats. *Journal of Social Behavior and Personality, 10*, 255-263.
- Lushene, R. E., O'Neil, H. F., & Dunn, T. (1974). Equivalent validity of a completely computerized MMPI. *Journal of Personality Assessment, 38*, 353-361.
- Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology, 74*, 72-80.
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*, 449-458.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.
- Millstein, S. G. (1987). Acceptability and reliability of sensitive information collected via computer interview. *Educational and Psychological Measurement, 47*, 523-533.
- Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). Measurement of organizational commitment. *Journal of Vocational Behavior, 14*, 224-247.
- Niehoff, B. P., & Moorman, R. H. (1993). Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. *Academy of Management Journal, 36*, 527-556.
- Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology, 56*, 733-752.
- Potosky, D., & Bobko, P. (1997). Computer versus paper-and-pencil administration mode and response distortion in noncognitive selection tests. *Journal of Applied Psychology, 82*, 293-299.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-566.
- Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*, 754-775.
- Riordan, C. M., & Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management, 20*, 643-671.
- Rosenfeld, P., Doherty, L. M., & Carroll, L. (1987). Microcomputer-based organizational survey assessment: Applications to training. *Journal of Business and Psychology, 2*, 182-193.
- Rosenfeld, P., Doherty, L. M., Vicino, S. M., Kantor, J., & Greaves, J. (1989). Attitude assessment in organizations—Testing three microcomputer-based survey-systems. *Journal of General Psychology, 116*, 145-154.
- Rosenfeld, P., Giacalone, R. A., Knouse, S. B., Doherty, L. M., Vicino, S. M., Kantor, J., et al. (1991). Impression management, candor, and microcomputer-based organizational surveys—an individual-differences approach. *Computers in Human Behavior, 7*, 23-32.
- Ryan, A. M., Chan, D., Ployhart, R. E., & Slade, L. A. (1999). Employee attitude surveys in a multinational organization: Considering language and culture in assessing measurement equivalence. *Personnel Psychology, 52*, 37-58.
- Schulderberg, D. (1988). The MMPI is less sensitive to the automated testing format than it is to repeated testing: Item and scale effects. *Computers in Human Behavior, 4*, 285-298.
- Shadish, W. R., Cook, T. D., & Campbell, T. L. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
- Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969). *The measurement of satisfaction in work and retirement: A strategy for the study of attitudes*. Oxford, UK: Rand McNally.
- Stanton, J. M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology, 51*, 709-725.
- Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.

- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.
- Synodinos, N. E., & Brennan, J. M. (1990). Evaluating microcomputer interactive survey software. *Journal of Business and Psychology*, 4, 483-493.
- Synodinos, N. E., Papacostas, C. S., & Okimoto, G. M. (1994). Computer-administered versus paper-and-pencil surveys and the effect of sample selection. *Behavior Research Methods, Instruments, & Computers*, 26, 395-401.
- Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, 280, 867-873.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.
- van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13, 29-37.
- Vansickle, T. R., Kimmel, C., & Kapes, J. T. (1989). Test retest equivalency of the computer-based and paper pencil versions of the Strong-Campbell Interest Inventory. *Measurement and Evaluation in Counseling and Development*, 22, 88-93.
- Weiss, D. J., Dawis, R. V., England, G. W., & Lofquist, L. H. (1967). *Manual for the Minnesota Satisfaction Questionnaire*. Minneapolis: University of Minnesota Press.
- White, D. M., Clements, C. B., & Fowler, R. D. (1985). A comparison of computer administration with standard administration of the MMPI. *Computers in Human Behavior*, 1, 153-162.
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, 9, 1-26.
- Wilkerson, J. M., Nagao, D. H., & Martin, C. L. (2002). Socially desirable responding in computerized questionnaires: When questionnaire purpose matters more than the mode. *Journal of Applied Social Psychology*, 32, 544-559.
- Wilson, F. R., Genco, K. T., & Yager, G. G. (1985). Assessing the equivalence of paper-and-pencil vs. computerized tests: Demonstration of a promising methodology. *Computers in Human Behavior*, 1, 265-275.